Genome-Wide Analysis of 18 Epstein-Barr Viruses Isolated from Primary Nasopharyngeal Carcinoma Biopsy Specimens

Chaofeng Tu,^{a,b,c} Zhaoyang Zeng,^{a,b,c} Peng Qi,^b Xiayu Li,^c Zhengyuan Yu,^b Can Guo,^{a,b,c} Fang Xiong,^a Bo Xiang,^{a,b,c} Ming Zhou,^{a,b,c} Zhaojian Gong,^b Qianjin Liao,^{b,d} Jianjun Yu,^d Yi He,^{b,d} Wenling Zhang,^b Xiaoling Li,^{a,b,c} Yong Li,^{b,e} Guiyuan Li,^{a,b,c} Wei Xiong^{a,b,c}

The Key Laboratory of Carcinogenesis of the Chinese Ministry of Health, Xiangya Hospital, Central South University, Changsha, Hunan, China^a; The Key Laboratory of Carcinogenesis and Cancer Invasion of the Chinese Ministry of Education, Cancer Research Institute, Central South University, Changsha, Hunan, China^b; Hunan Key Laboratory of Nonresolving Inflammation and Cancer, Disease Genome Research Center, the Third Xiangya Hospital, Central South University, Changsha, Hunan, China^c; Department of Head and Neck Surgery, Hunan Cancer Hospital and the Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, Hunan, China^d; Department of Cancer Biology, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA^e

ABSTRACT Epstein-Barr virus (EBV) is a ubiquitous gammaherpesvirus that is highly prevalent in almost all human populations and is associated with many human cancers, such as nasopharyngeal carcinoma (NPC), Hodgkin's disease, and gastric carcinoma. However, in these EBV-associated cancers, only NPC exhibits remarkable ethnic and geographic distribution. We hypothesized that EBV genomic variations might contribute to the pathogenesis of different human cancers in different geographic areas. In this study, we collected 18 NPC biopsy specimens from the Hunan Province in southern China and *de novo* assembled 18 NPC biopsy specimen-derived EBV (NPC-EBV) genomes, designated HN1 to HN18. This was achieved through target enrichment of EBV DNA by hybridization, followed by next-generation sequencing, to reveal sequence diversity. These EBV genomes harbored 20,570 variations totally, including 20,328 substitutions, 88 insertions, and 154 deletions, compared to the EBV reference genome. Phylogenetic analysis revealed that all NPC-EBV genomes were distinct from other EBV genomes. Furthermore, HN1 to HN18 had some nonsynonymous variations in EBV genes including genes encoding latent, early lytic, and tegument proteins, such as substitutions within transmembrane domains 1 and 3 of LMP1, FoP_duplication, and zf-AD domains of ENBA1, in addition to aberrations in noncoding regions, especially in BamHI A rightward transcript microRNAs. These variations might have potential biological significance. In conclusion, we reported a genome-wide view of sequence variation in EBV isolated from primary NPC biopsy specimens obtained from the Hunan Province. This might contribute to further understanding of how genomic variations contribute to carcinogenesis, which would impact the treatment of EBV-associated cancer.

IMPORTANCE Nasopharyngeal carcinoma (NPC) is highly associated with Epstein-Barr virus (EBV) infection and exhibits remarkable ethnic and geographic distribution. Hunan Province in southern China has a high incidence rate of NPCs. Here, we report 18 novel EBV genome sequences from viruses isolated from primary NPC biopsy specimens in this region, revealing whole-genome sequence diversity.

KEYWORDS nasopharyngeal carcinoma, NPC, genetic variations, virus capture sequencing, phylogenetic analysis, Epstein-Barr virus, EBV

Received 23 February 2017 Accepted 8 May 2017

Accepted manuscript posted online 21 June 2017

Citation Tu C, Zeng Z, Qi P, Li X, Yu Z, Guo C, Xiong F, Xiang B, Zhou M, Gong Z, Liao Q, Yu J, He Y, Zhang W, Li X, Li Y, Li G, Xiong W. 2017. Genome-wide analysis of 18 Epstein-Barr viruses isolated from primary nasopharyngeal carcinoma biopsy specimens. *J Virol* 91:e00301-17. <https://doi.org/10.1128/JVI.00301-17>.

Editor Rozanne M. Sandri-Goldin, University of California, Irvine

Copyright © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Guiyuan Li, lgy@csu.edu.cn, or Wei Xiong, xiongwei@csu.edu.cn.

C.T. and Z.Z. contributed equally to this article.

Epstein-Barr virus (EBV) is a ubiquitous gammaherpesvirus that is highly prevalent in all human populations (1). It is associated with both nonmalignant diseases such as infectious mononucleosis (2) and a number of human cancers such as undifferentiated nasopharyngeal carcinoma (NPC) (3), Hodgkin's disease (4), and gastric carcinoma (5). However, regarding EBV-associated cancers, only NPC exhibits remarkable ethnic and geographic distribution, with a particularly high prevalence in southern China and Southeast Asia (3). Hunan Province represents a region with a high NPC prevalence in southern China (6–8). We hypothesized that EBV genomic variations might contribute to the pathogenesis of different human cancers in different geographical locations. However, genome-wide variation analysis of EBV subtypes in Hunan Province has not yet been performed.

Traditional studies on the relationship between genetic variations in EBV and their contribution to the pathogenesis of NPC have been performed using a small subset of EBV genes such as *LMP1*, *BHRF1*, *BZLF1*, and *EBNA1* (9–17). However, genetic variation analysis based only on a small portion of the EBV genome is not sufficient to assess the precise relationship between genome variations and the pathogenesis of human cancers. Therefore, it is urgent to identify EBV genetic variations using the whole EBV genome. With the development of traditional Sanger sequencing and next-generation sequencing (NGS) technology, 28 genomes from EBVs isolated from several diseases have been reported, including B95-8 (18), AG876 (19), Akata (20), Mutu (20), GD1 (21), GD2 (22), HKNPC1-9 (23, 24), C666-1 (25, 26), Raji (27), K4413-Mi (28), K4123-Mi (28), and EBVaGC1-9 (29). These sequencing studies all revealed that EBV subtypes from different regions exhibit significant regional distribution.

In this study, EBV genomes from 18 NPC biopsy specimens collected from Hunan Province, southern China, were *de novo* assembled through target hybridization enrichment, followed by NGS, which is analogous to human exome sequencing. The sequences of 18 NPC biopsy-specimen-derived EBV (NPC-EBV) genomes were designated HN1 to HN18. Heterogeneity, variations in protein-coding and noncoding regions, and phylogenetic and protein sequence variation analyses were subsequently performed to assess the genomic diversity of NPC-EBV genomes. These data are expected to help clarify the complex role that NPC-EBV genomes play in the pathogenesis of NPC.

RESULTS

Target enrichment, sequencing, and *de novo* assembly of EBV genomes. In this study, 18 EBV-positive NPC biopsy specimens were collected from Hunan Province in southern China and subjected to capture sequencing. Initially, an average total of 10,000,000 reads (1 gigabase of raw data) were collected from each sample. Subsequently, all sequencing reads from 18 NPC-derived EBV genomes were mapped to the EBV reference sequence (accession number [NC007605](#)). The reads that matched the EBV reference sequence were used for EBV genome assembly using Burrows-Wheeler Aligner (BWA) programs (29). All 18 NPC-EBV genomes, designated HN1 to HN18, were successfully sequenced and assembled. The coverage of HN1 to HN18, mapped to the reference EBV sequence, ranged from 98.93 to 99.96%. The average sequencing depth on target for these 18 NPC-EBV genomes was 4,047-fold, ranging from 248- to 11,751-fold, with assembled EBV genome sizes ranging from 168,444 bp (HN4) to 171,822 bp (HN15). Details of the sequencing data are listed in Table 1.

Variation analysis of NPC-EBV genomes. A total of 20,570 variations were observed in these 18 EBV genomes compared to the reference EBV sequence, including 20,238 substitutions, 88 insertions, and 154 deletions. Among these variations, 15,440 substitutions, 3 insertions, and 8 deletions were located in the coding regions of the EBV genomes, with a median variation rate of 0.502% and a range of 0.458% (HN10) to 0.583% (HN2) (Table 2). Figure 1 illustrates the variations in the NPC-EBV genomes (HN1 to HN9 [Fig. 1A] and HN10 to HN18 [Fig. 1B]) relative to the reference EBV genome. Compared to the reference EBV genome, the highest incidence of variations was found at kb 95 to 105 of the EBV genome, corresponding to the location of the genes *EBNA1*

TABLE 1 Summary of data from the Illumina HiSeq 2000 analysis

Sample	No. of reads		Mapped bases (bp)	Coverage (%)	Sequencing depth (X)	Length (bp)	GenBank accession no.
	Raw	Mapped					
HN1	16,726,133	8,929,794	1,339,469,100	99.95	7,799.72	171,499	AB850643
HN2	22,748,186	13,455,525	2,018,328,750	99.96	11,751.55	171,751	AB850644
HN3	12,984,099	6,552,481	982,872,150	99.94	5,723.72	171,672	AB850648
HN4	11,537,026	5,946,605	891,990,750	99.93	5,194.87	168,444	AB850649
HN5	3,463,466	1,121,847	168,277,050	99.79	981.45	171,733	AB850652
HN6	12,005,666	4,861,088	729,163,200	99.96	4,245.57	171,054	AB850653
HN7	4,059,335	1,782,358	267,353,700	99.67	1,561.16	171,548	AB850655
HN8	15,751,237	7,735,218	1,160,282,700	99.82	6,764.87	169,829	AB850657
HN9	7,517,402	2,921,033	438,154,950	99.81	2,554.88	171,344	AB850659
HN10	7,406,538	2,278,723	341,808,450	99.92	1,991.00	169,720	AB850645
HN11	10,870,010	4,688,792	703,318,800	99.92	4,096.73	168,509	AB850646
HN12	6,442,145	898,244	134,736,600	99.57	787.57	171,663	AB850647
HN13	5,752,790	2,863,733	429,559,950	99.94	2,501.44	171,440	AB850650
HN14	13,596,811	6,987,188	1,048,078,200	99.92	6,104.69	171,518	AB850651
HN15	5,208,005	281,399	42,209,850	98.93	248.32	171,822	AB850654
HN16	16,622,886	6,137,085	920,562,750	99.95	5,360.15	171,609	AB850656
HN17	8,094,712	3,686,012	552,901,800	99.93	3,220.14	171,394	AB850658
HN18	6,138,550	2,257,724	338,658,600	99.90	1,972.99	170,059	AB850660

(encoding EBNA1 and related to replication of the EBV genome) and *BKRF1* (which is associated with host infection). The second highest incidence of variations was found at kb 165 to 172, corresponding to the location of the *LMP1* gene (encoding LMP1, which is known as an oncogenic virus protein) (Fig. 1C).

The EBV genome encodes 86 proteins (30). To investigate genetic variations and their potential roles in the carcinogenesis of NPC, we classified affected proteins and found that nonsynonymous variations in latent membrane protein-encoding genes were of the highest proportion (21.4%), based on all NPC-EBV genome variations, with an average of 120 nonsynonymous substitutions per sample. Variations in tegument and membrane protein-encoding genes, respectively, accounted for 19.6 and 12.0% of mutations (Fig. 2 and Table 3), supporting the important role of latent EBV infection in NPC oncogenesis.

TABLE 2 Variations in HN1-HN18 in comparison to the reference Epstein-Barr virus genome

		No. of variations ^a						
Sample	Length (bp)	Substitutions		Insertions		Deletions		Variation rate (%)
		CRs	NCRs	CRs	NCRs	CRs	NCRs	
HN1	171,499	940	288	0	3	1	5	0.548
HN2	171,751	1001	297	0	1	1	5	0.583
HN3	171,672	840	273	0	3	1	3	0.489
HN4	168,444	841	231	0	2	0	5	0.499
HN5	171,733	797	286	2	7	0	12	0.464
HN6	171,054	886	246	0	3	0	6	0.518
HN7	171,548	798	272	0	7	0	9	0.465
HN8	169,829	812	279	0	3	0	5	0.478
HN9	171,344	788	265	0	3	0	8	0.460
HN10	169,720	777	273	0	5	1	7	0.458
HN11	168,509	938	236	0	4	0	9	0.557
HN12	171,663	863	260	0	9	1	12	0.503
HN13	171,440	838	263	0	4	0	3	0.489
HN14	171,518	840	279	0	3	0	5	0.490
HN15	171,822	859	304	0	14	1	31	0.500
HN16	171,609	833	298	0	6	0	7	0.485
HN17	171,394	891	265	1	2	1	7	0.520
HN18	170,059	898	273	0	6	1	7	0.528

^aCRs, coding regions; NCRs, noncoding regions.

FIG 1 Distribution of nonsynonymous variations in 18 NPC biopsy-derived Epstein-Barr virus (NPC-EBV) genomes. Genome sequences of 18 NPC-EBVs, designated HN1 to HN18, were assembled and compared to the reference EBV sequence (NC007605). Circos plots representing the nonsynonymous variations in HN1 to HN9 (A) and HN10 to HN18 (B) tracks from outer to inner represent the following: track 1 (outermost), coding exons for latent genes (red), lytic genes (blue), and RPS1/miRNAs (green); track 2, repeat and regulatory elements (black); and track 3, nonsynonymous variations (bright green) in the NPC-EBV genomes compared to the EBV reference genome sequence. (C) Genome-wide distribution of single-nucleotide variation (SNVs) in HN1 to HN18. The genomic positions are related to the EBV reference sequence (NC007605). SNVs were identified by cross_match and BWA software. The highest incidences of variations were found at 95 to 105 kb (where the *EBNA1* and *BKRF1* genes are located) and 165 to 172 kb (which includes the *LMP1* gene). (See File S1 in the supplemental material for an enlargeable version of the figure.)

Phylogenetic analysis of NPC-EBV genomes. Phylogenetic trees representing whole NPC-EBV genomes were constructed from 18 sequences assembled in this study and other published EBV genomes (18–29) (Fig. 3). All NPC-EBV genomes from Hunan Province (HN1 to HN18) clustered in one branch, which was separate from EBV subtypes from other areas, indicating that EBV subtypes show specific geographical distribution.

Variations in EBV protein-encoding genes and noncoding RNA. Phylogenetic analysis of individual gene sequences that demonstrated higher incidences of variation, including *LMP1*, *BZLF1*, *EBNA1*, *BKRF2*, *BKRF3*, and *BKRF4*, demonstrated similar results compared to whole EBV genome sequence analysis (Fig. 4); specifically, HN1 to HN18 clustered at one branch. Based on the amino acid sequences of these genes, we found

FIG 2 Categories of nonsynonymous variations in 18 NPC-EBVs. Nonsynonymous variations in the HN1 to HN18 samples were categorized based on the functions of EBV-encoded proteins. Most amino acid changes were located in latency-associated proteins (red), followed by tegument-associated proteins (light blue) and membrane-associated proteins (green).

that *LMP1* has specific regions containing substitutions, for example, at positions 31 to 34 within transmembrane domain 1 and at positions 91 to 94 within transmembrane domain 3 in HN1 to HN18 (Fig. 5A). These variations might presumably influence LMP1 signaling functions. Furthermore, HN1 to HN18 had two regions that demonstrated a high incidence of substitution polymorphism within the EBNA1 protein FoP_duplication domain (residues 16 [Q/E] and 18 [E/G]) and zf-AD domain (residues 543 [P/S] and

TABLE 3 Summary of nonsynonymous variations in Epstein-Barr virus coding genes

Sample	No. of variations									Total
	Latency	Tegument	Membrane	Replication	Transcription factor	Capsid	Packaging	Nucleotide metabolism	Unknown	
HN1	167	123	94	77	59	52	28	31	96	727
HN2	165	132	112	72	74	74	32	44	82	787
HN3	107	115	75	69	39	62	21	23	78	589
HN4	152	150	55	37	41	50	18	20	58	581
HN5	125	101	63	53	32	47	23	19	84	547
HN6	131	124	75	61	43	56	20	26	91	627
HN7	122	101	60	51	29	50	25	19	78	535
HN8	121	104	65	74	31	56	28	16	70	565
HN9	97	97	51	44	24	44	24	14	75	470
HN10	111	85	57	40	19	30	22	15	48	427
HN11	106	108	72	62	40	51	19	39	48	545
HN12	61	133	76	47	45	61	15	36	51	525
HN13	118	126	80	63	33	64	22	30	79	615
HN14	127	94	62	57	47	52	31	19	71	560
HN15	92	100	66	54	41	50	12	22	53	490
HN16	109	93	43	54	39	47	27	17	64	493
HN17	146	91	64	42	37	52	18	37	76	563
HN18	113	106	54	72	31	50	16	30	42	514

FIG 3 Phylogenetic analyses of 18 EBV genomes assembled in this study and other EBV strains based on whole EBV genomic sequences. The numbers at the internal nodes correspond to bootstrap values obtained based on the analysis of 500 replicates.

544 [Q/Y]) compared to that of other subtypes (Fig. 5B). In addition, HN1 to HN18 was shown to have one region of substitution within the glycop_{gl} domain (residues 38 [H/R] and 81 [N/H]) of *BKRF2* (Fig. 5C) and in the UDG domain (residues 77 to 84) of *BKRF3* (Fig. 5D).

It has been reported that EBV encodes 44 mature microRNAs that are grouped in two clusters located surrounding the *BHRF1* gene (*BHRF1* cluster) and within the BamHI A rightward transcript (*BART* cluster). Therefore, variation analysis was also performed on these noncoding regions of the EBV genomes. One substitution was found in 15 of 18 NPC-EBV genomes, which was in the mature EBV miRNA *EBV-miR-BART19-5p* (34 T/Cs on the stem-loop sequence of pre-*miR-BART19*; Fig. 6A); this was also found in NPC-EBV genomes from southern China, Guangdong (GD2), and Hong Kong (HKNPC1-9) (22–24). In addition, substitution variations, namely, 47 T/Cs on the stem-loop sequence of pre-*miR-BART19* (Fig. 6A), 8 G/As on pre-*miR-BART10* (Fig. 6B), 93 C/Ts on pre-*miR-BART17* (Fig. 6C) and 7 C/Ts on pre-*miR-BART8* (Fig. 6D) were also found within the loop, 3' end, or 5' end of these pre-EBV-miRNAs, respectively, but were not found to be located in the mature miRNA sequences. These variations could influence the mature processing and expression level of these EBV-encoded miRNAs.

DISCUSSION

NPC is a squamous-cell carcinoma that arises in the epithelial lining of the nasopharynx. This neoplasm has notable ethnic and geographic distribution, exhibiting high prevalence in southern China but low incidence in other parts of the world (31). Genetic susceptibility and EBV infection are important etiological factors of NPC (32–44). However, how environmental (for example, EBV) (45) and genetic risk factors (46–58) interact during the carcinogenesis of NPC is still not well understood.

Previous studies have demonstrated associations between host genetic variations

FIG 4 Phylogenetic analysis of some important EBV coding genes. Sequences of *LMP1* (A), *BZLF1* (B), *EBNA1* (C), *BKRF2* (D), *BKRF3* (E), and *BKRF4* (F) genes derived from this study and those from previously reported EBV genomes were analyzed. The numbers at the internal nodes correspond to the bootstrap values obtained based on the analysis of 500 replicates.

FIG 5 Sequence variations in the amino acid sequences of important EBV proteins. Alignment of partial protein sequences of LMP1 (A), EBNA1 (B), BKRf2 (C), and BKRf3 (D) based on 18 EBV genomes assembled in this study and those of other EBV strains. Variant positions are marked in yellow. Functional domains in this protein are indicated above the consensus sequence. (See File S1 in the supplemental material for an enlargeable version of the figure.)

FIG 6 Variations in EBV-encoded miRNAs. Alignment of pre-miRNA sequences of EBV-encoded miRNAs BART19 (A), BART10 (B), BART17 (C), and BART8 (D) from 18 EBV genomes assembled in this study and those of other EBV strains. Variant positions are marked in red; mature miRNA sequences are marked above the consensus sequence. (See File S1 in the supplemental material for an enlargeable version of the figure.)

and certain EBV variations and NPC risk (59). Linkage analysis is a major strategy to identify genes that contribute to NPC risk. However, four linkage studies have been performed, wherein susceptibility loci, including 6p21 (59), 4p15.1-q12 (60), 3p21.31-21.2 (6, 7), and 5p13 (61), were implicated as NPC risk factors in families of southern Chinese origin such as those from Guangdong, Hong Kong, and Hunan. These results are not concordant, since no studies provided supporting linkage evidence for each other. This might be partially explained by the heterogeneity of the population or NPC subtype.

EBV is highly prevalent in all human populations and is associated with both nonmalignant disease and some human cancers (32); however, only NPC has remarkable ethnic and geographic distribution, indicating that the EBV genome might also exhibit heterogeneity with different subtypes resulting in different diseases (62, 63). We hypothesized that the subtype of EBV that is prevalent in southern China might have unique genetic variations that play an important role in the development of NPC as opposed to other diseases. Actually, NPC-EBV strains have been characterized by genotyping polymorphic markers in the *EBER1*, *EBER2*, *LMP1*, *BHRF1*, *BZLF1*, and *EBNA1* gene loci; however, genetic variations in the small subsets of genes investigated were not sufficient to assess the geographical distribution of EBV variations and their precise association with disease. NGS technology provides an important and highly efficient method to delineate the genomic sequences of EBV genomes, and as such, several EBV

strains derived from Guangdong (GD1 and GD2) and Hong Kong (HKNPC1 to HKNPC9) have been sequenced and assembled (21–24). We previously reported that the susceptibility locus on chromosome 3p21 in NPC families from Hunan Province (6) is distinct from other susceptibility loci in NPC families from Guangdong (60) and Hong Kong (61); therefore, determining EBV genomic sequences from NPC patients in Hunan Province is worthwhile.

In this study, we sequenced and assembled 18 EBV genomes from NPC patients in Hunan Province through target hybridization enrichment, followed by an NGS strategy, a technique that is associated with better specificity, lower cost, and a higher depth of coverage (the average sequencing depth was 4,047-fold). This suggests that our results are more accurate and credible. A total of 20,570 variations were observed in these 18 NPC-EBV genomes compared to the EBV reference genome, and the average number of variations in protein coding regions was 261 per EBV genome. Among them, latent genes such as *LMP1*, *EBNA1*, and *BKRF2* represented nearly one-third of all variations. The EBV reference genome was constructed using B95-8 (infecting marmoset B cells) as the backbone, with an 11-kb deletion segment provided by the Raji (Burkitt's lymphoma cell line) sequence; this suggests that latent EBV infection plays a specific role in NPC tumorigenesis and that this role might be different from that in other human cancers, especially B-cell-lineage malignancies. Actually, there are four different latent infection statuses (latencies I to IV) of EBV in different EBV-associated human cancers, and EBV expresses different latency genes; for example, Burkitt's lymphoma is latency I, and nasopharyngeal carcinoma is latency II.

In this study, we also performed phylogenetic analysis according to whole EBV genomic sequences and based on certain important EBV genes such as *LMP1*, *EBNA1*, and *BKRF2*. Results showed that EBV subtypes in Hunan areas could be distinguished from those of other areas. In addition, EBV subtypes in Southeast Asia and North Africa were shown to cluster as one branch that is separate from subtypes from other areas; this was based on *LMP1*, *EBNA1*, and *BKRF2* gene sequences, which indicates that *LMP1*, *EBNA1*, and *BKRF2* can act as potential classification biomarkers for EBV subtypes. Based on *BKRF3*, *BKRF4*, and *BZLF1*, the EBV subtypes in Hunan areas also clustered as one branch, whereas HKNPC6 and HKNPC7 from Hong Kong and Mutu, as well as K4123-Mi from North Africa, clustered as another branch, indicating that subtypes can be differentiated based only on a single gene. Clinical information showed that HKNPC6 and HKNPC7 were stage IV and that the differentiated degree of these two samples was higher than that of other samples, indicating that clinical stages might affect HKNPC6 and HKNPC7 clustering. By analyzing the *LMP1*-encoding amino acid sequence, insertions/deletions within the transmembrane domain and Pfam domain of *LMP1* correlated with geographical distribution of the samples, which was confirmed by phylogenetic analysis, indicating that *LMP1* could be a potential biomarker for the classification of various subtypes of EBV.

EBV encodes 44 microRNAs (miRNAs), including the BART cluster and BHRF cluster (64). The largest set is the BART miRNAs, which consists of 40 miRNAs that are expressed in all forms during EBV infection, promoting viral latency or cancer development by targeting both viral and cellular genes (65–67). Many studies have reported that polymorphisms in human miRNAs or genetic variations in miRNA-binding sites in the 3' untranslated region (UTR) of human genes influence cancer susceptibility or promote cancer progression, affecting prognosis (68). In this study, several variations in EBV miRNAs were found; especially, polymorphisms in the mature miRNA *EBV-miR-BART19-5p* might influence minimum free energy or potency and stability during binding to the 3' UTR of target genes. This could alter host target gene expression and affect cancer progression. Other variations were identified in pre-miRNA sequences, albeit not in EBV-miRNA mature sequences, but rather located in stem-loop sequences. These variations might lead to structural alterations in pre-miRNAs and could result in the aberrant expression of EBV-miRNAs. The detailed biological function of these EBV mRNA variants is worthy of further study.

TABLE 4 Clinical features of nasopharyngeal carcinoma biopsy specimens

Case	Gender	Age (yr)	TNM stage	Overall stage
HN1	Male	59	T3N2M0	III
HN2	Male	61	T4N3M0	IVA
HN3	Female	54	T3N2M0	III
HN4	Female	51	T4N3M0	IVA
HN5	Female	28	T2N3M0	IVA
HN6	Male	38	T2N2M1	IVB
HN7	Male	70	T3N2M0	III
HN8	Female	49	T4N3M0	IVA
HN9	Male	36	T4N2M0	IVA
HN10	Male	68	T4N3M0	IVA
HN11	Male	44	T4N2M0	IVA
HN12	Male	47	T4N2M0	IVA
HN13	Female	67	T2N2M0	III
HN14	Female	72	T2N2M1	IVB
HN15	Male	52	T3N0M0	III
HN16	Female	63	T4N2M0	IVA
HN17	Male	54	T4N3M0	IVA
HN18	Female	50	T4N2M1	IVB

In summary, we performed virus sequence capture sequencing to sequence 18 EBV-associated NPC biopsy specimens from the Hunan area. These EBV genomic sequences provide an important reference to further identify variations in EBV. These data will provide a more appropriate standard to classify EBV subtypes and provide a large amount of resources to further define the epidemiological and pathogenic roles of EBV in associated diseases.

MATERIALS AND METHODS

Sample DNA preparation. A total of 18 primary undifferentiated NPC biopsy specimens were collected from the Hunan Cancer Hospital in Changsha, Hunan Province, China. Ethics approval was obtained from the Institutional Review Board. Detailed clinical information about the patients is summarized in the Table 4. DNA extraction was performed using a Qiagen blood and tissue kit according to the manufacturer's protocol (Qiagen, Hilden, Germany) (69–71). DNA concentrations were determined using a Nano-Drop (Thermo Fisher Scientific, Wilmington, DE).

Workflow for capture sequencing of EBV genomes. In brief, the workflow for capture sequencing of EBV genomes included EBV capturing probe preparation, NPC biopsy DNA library preparation, target capture and index tagging, sequencing, *de novo* assembly, and bioinformatics analysis. A flow chart for the complete workflow is shown in Fig. 7. These steps are described in further detail as follows.

EBV genome capture and sequencing. EBV probes were designed according to the EBV full-length reference genomes by MyGenostics (Beijing, China). Biotin-labeled deoxynucleotide triphosphates were used during amplification of the EBV genome sequences. The overall experiment was conducted according to the manufacturer's protocol. In brief, products were purified using Ampure beads (Beckman Coulter, Inc.) and a QIAEX II gel extraction kit (Qiagen) and fragmented randomly to 250- to 500-bp

FIG 7 Workflow for sequencing and analysis of EBV genomes from NPC biopsy specimens.

segments using a Covaris E-210 ultrasonicator (Covaris, Inc., Woburn, MA). Subsequently, the single EBV capture probe was denatured at 94°C.

Purified genomic DNA (containing human and EBV genomes) from NPC biopsy specimens was fragmented randomly to produce a 170- to 200-bp DNA library using an ultrasonicator Covaris E-210 (70, 72). DNA fragments were end repaired using a DNA terminator end repair kit (Lucigen, Middleton, WI) and purified using a QIAquick PCR purification kit (Qiagen) according to the manufacturer's protocol. Subsequently, 3'-dA tailing was performed using a Klenow fragment (Enzymatics, Beverly, MA). After 10 cycles of PCR amplification, the dA-tailed DNA fragments were purified using 2% polyacrylamide gels (Life Technologies, Carlsbad, CA), and the DNA library was analyzed using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA).

EBV capturing probes and DNA libraries, purified from NPC biopsy specimens, were hybridized at 47°C for 24 h, after which nonhybridized DNA fragments were cleared. The purified fragments were amplified for 16 cycles using AccuPrime Pfx DNA polymerase (Life Technologies), and the sequence library was used for next-generation sequencing. Cluster generation and 150-bp paired-end sequencing reactions were performed using an Illumina HiSeq 2000 sequencer (Illumina, Inc., San Diego, CA) using 101 cycles and in accordance with the manufacturer's instructions.

De novo assembly of EBV genomes. All sequencing reads were *de novo* assembled using SOAPdenovo software (73), which merged overlapping reads based on the de Bruijn graph algorithm to generate the contigs. The paired-end data were then used to link contigs into scaffolds, and the assembled scaffolds were aligned to the EBV reference genome to fill the gaps, which were mostly located in the repeat sequence regions.

Detection of SNVs. All sequencing reads were aligned to the EBV reference genome using BWA software. Single-nucleotide variations (SNVs) were detected by SAMtools software (74) and applying the following criteria to filter the raw variation results: (i) the minimal depth cover of the variation site was no less than 4; (ii) the minimal variation quality score was not less than 20; (iii) the distance between two nearby variations was not less than 5; and (iv) the reads supporting the mutated allele were not fewer than 4.

Phylogenetic analysis. MAFFT software (75) was used to align HN1 to HN18 EBV subtypes with published EBV genomes, including the EBV reference genome (NC007605), B95-8 (V01555), AG876 (DQ279927), Akata (KC207813), Mutu (KC207814), K4123-Mi (KC440851), K4413-Mi (KC440852), GD1 (AY961628), GD2 (HQ020558), HKNPC1 (JQ009376), HKNPC2 to HKNPC9 (KF992564 to KF992571), C666-1 (KC617875), and Raji (KF717093). Molecular Evolutionary Genetics Analysis (MEGA) software (76) was used to perform the phylogenetic analysis using the neighbor-joining algorithm, based on multiple sequence alignments of all EBV genomes and their encoded genes. Protein sequences were aligned with MegAlign (DNASTar-Lasergene) software.

Accession number(s). Sequence data for the 18 NPC-EBV genomes, HN1 to HN18, were submitted to the DDBJ and GenBank database under accession numbers AB850643 to AB850660 (Table 1).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JVI.00301-17>.

SUPPLEMENTAL FILE 1, PDF file, 7.3 MB.

ACKNOWLEDGMENTS

This study was supported in part by grants from the National High Technology Research and Development Program of China (2012AA02A206), the National Natural Science Foundation of China (81372907, 81301757, 81472531, 81402009, 81572787, 81672683, and 81672993), the Programme of Introducing Talents of Discipline to Universities (111-2-12), the Natural Science Foundation of Hunan Province (13JJ3039 and 2015JJ1022), and the Fundamental Research Funds for the Central Universities of Central South University (2016zzts478).

The authors declare that there are no conflicts of interest regarding this work.

REFERENCES

- Henle W, Henle G, Zajac BA, Pearson G, Waubke R, Scriba M. 1970. Differential reactivity of human sera with early antigens induced by Epstein-Barr virus. *Science* 169:188–190. <https://doi.org/10.1126/science.169.3941.188>.
- De Waele M, Thielemans C, Van Camp BK. 1981. Characterization of immunoregulatory T cells in EBV-induced infectious mononucleosis by monoclonal antibodies. *N Engl J Med* 304:460–462. <https://doi.org/10.1056/NEJM198102193040804>.
- Wei WI, Sham JS. 2005. Nasopharyngeal carcinoma. *Lancet* 365: 2041–2054. [https://doi.org/10.1016/S0140-6736\(05\)66698-6](https://doi.org/10.1016/S0140-6736(05)66698-6).
- Merli F, Luminari S, Gobbi PG, Cascavilla N, Mammi C, Ilariucci F, Stelitano C, Musso M, Baldini L, Galimberti S, Angrilli F, Polimeno G, Scalzulli PR, Ferrari A, Marcheselli L, Federico M. 2016. Long-term results of the HD2000 trial comparing ABVD versus BEACOPP versus COPP-EBV-CAD in untreated patients with advanced Hodgkin lymphoma: a study by Fondazione Italiana Linfomi. *J Clin Oncol* 34:1175–1181. <https://doi.org/10.1200/JCO.2015.62.4817>.
- Hino R, Uozaki H, Inoue Y, Shintani Y, Ushiku T, Sakatani T, Takada K, Fukayama M. 2008. Survival advantage of EBV-associated gastric carcinoma: survivin up-regulation by viral latent membrane protein 2A. *Cancer Res* 68:1427–1435. <https://doi.org/10.1158/0008-5472.CAN-07-3027>.
- Xiong W, Zeng ZY, Xia JH, Xia K, Shen SR, Li XL, Hu DX, Tan C, Xiang JJ,

- Zhou J, Deng H, Fan SQ, Li WF, Wang R, Zhou M, Zhu SG, Lu HB, Qian J, Zhang BC, Wang JR, Ma J, Xiao BY, Huang H, Zhang QH, Zhou YH, Luo XM, Zhou HD, Yang YX, Dai HP, Feng GY, Pan Q, Wu LQ, He L, Li GY. 2004. A susceptibility locus at chromosome 3p21 linked to familial nasopharyngeal carcinoma. *Cancer Res* 64:1972–1974. <https://doi.org/10.1158/0008-5472.CAN-03-3253>.
7. Zeng Z, Zhou Y, Zhang W, Li X, Xiong W, Liu H, Fan S, Qian J, Wang L, Li Z, Shen S, Li G. 2006. Family-based association analysis validates chromosome 3p21 as a putative nasopharyngeal carcinoma susceptibility locus. *Genet Med* 8:156–160. <https://doi.org/10.1097/01.gim.0000196821.87655.d0>.
 8. Zeng Z, Huang H, Zhang W, Xiang B, Zhou M, Zhou Y, Ma J, Yi M, Li X, Li X, Xiong W, Li G. 2011. Nasopharyngeal carcinoma: advances in genomics and molecular genetics. *Sci China Life Sci* 54:966–975. <https://doi.org/10.1007/s11427-011-4223-5>.
 9. Grunewald V, Bonnet M, Boutin S, Yip T, Louzir H, Levrero M, Seigneurin JM, Raphael M, Touitou R, Martel-Renoir D, Cochet C, Durandy A, Andre P, Lau W, Zeng Y, Joab I. 1998. Amino-acid change in the Epstein-Barr virus ZEBRA protein in undifferentiated nasopharyngeal carcinomas from Europe and North Africa. *Int J Cancer* 75:497–503. [https://doi.org/10.1002/\(SICI\)1097-0215\(19980209\)75:4<497::AID-IJC2>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0215(19980209)75:4<497::AID-IJC2>3.0.CO;2-O).
 10. Sacaze C, Henry S, Icart J, Mariame B. 2001. Tissue specific distribution of Epstein-Barr virus (EBV) BZLF1 gene variants in nasopharyngeal carcinoma (NPC) bearing patients. *Virus Res* 81:133–142. [https://doi.org/10.1016/S0168-1702\(01\)00376-8](https://doi.org/10.1016/S0168-1702(01)00376-8).
 11. Dardari R, Khyatti M, Cordeiro P, Odda M, El Gueddari B, Hassar M, Menezes J. 2006. High frequency of latent membrane protein-1 30-bp deletion variant with specific single mutations in Epstein-Barr virus-associated nasopharyngeal carcinoma in Moroccan patients. *Int J Cancer* 118:1977–1983. <https://doi.org/10.1002/ijc.21595>.
 12. Chang KP, Hao SP, Lin SY, Ueng SH, Pai PC, Tseng CK, Hsueh C, Hsieh MS, Yu JS, Tsang NM. 2006. The 30-bp deletion of Epstein-Barr virus latent membrane protein-1 gene has no effect in nasopharyngeal carcinoma. *Laryngoscope* 116:541–546. <https://doi.org/10.1097/01.mlg.0000201993.53410.40>.
 13. Nguyen-Van D, Ernberg I, Phan-Thi Phi P, Tran-Thi C, Hu L. 2008. Epstein-Barr virus genetic variation in Vietnamese patients with nasopharyngeal carcinoma: full-length analysis of LMP1. *Virus Genes* 37:273–281. <https://doi.org/10.1007/s11262-008-0262-9>.
 14. See HS, Yap YY, Yip WK, Seow HF. 2008. Epstein-Barr virus latent membrane protein-1 (LMP-1) 30-bp deletion and XhoI-loss is associated with type III nasopharyngeal carcinoma in Malaysia. *World J Surg Oncol* 6:18. <https://doi.org/10.1186/1477-7819-6-18>.
 15. Wang Y, Mo Y, Gong Z, Yang X, Yang M, Zhang S, Xiong F, Xiang B, Zhou M, Liao Q, Zhang W, Li X, Li X, Li Y, Li G, Zeng Z, Xiong W. 2017. Circular RNAs in human cancer. *Mol Cancer* 16:25. <https://doi.org/10.1186/s12943-017-0598-7>.
 16. Wang M, Zhao J, Zhang L, Wei F, Lian Y, Wu Y, Gong Z, Zhang S, Zhou J, Cao K, Li X, Xiong W, Li G, Zeng Z, Guo C. 2017. Role of tumor microenvironment in tumorigenesis. *J Cancer* 8:761–773. <https://doi.org/10.7150/jca.17648>.
 17. Tang Y, Wang J, Lian Y, Fan C, Zhang P, Wu Y, Li X, Xiong F, Li X, Li G, Xiong W, Zeng Z. 2017. Linking long noncoding RNAs and SWI/SNF complexes to chromatin remodeling in cancer. *Mol Cancer* 16:42. <https://doi.org/10.1186/s12943-017-0612-0>.
 18. Baer R, Bankier AT, Biggin MD, Deininger PL, Farrell PJ, Gibson TJ, Hatfull G, Hudson GS, Satchwell SC, Seguin C, et al. 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* 310:207–211. <https://doi.org/10.1038/310207a0>.
 19. Dolan A, Addison C, Gatherer D, Davison AJ, McGeoch DJ. 2006. The genome of Epstein-Barr virus type 2 strain AG876. *Virology* 350:164–170. <https://doi.org/10.1016/j.virol.2006.01.015>.
 20. Lin Z, Wang X, Strong MJ, Concha M, Baddoo M, Xu G, Baribault C, Fewell C, Hulme W, Hedges D, Taylor CM, Flemington EK. 2013. Whole-genome sequencing of the Akata and Mutu Epstein-Barr virus strains. *J Virol* 87:1172–1182. <https://doi.org/10.1128/JVI.02517-12>.
 21. Zeng MS, Li DJ, Liu QL, Song LB, Li MZ, Zhang RH, Yu XJ, Wang HM, Ernberg I, Zeng YX. 2005. Genomic sequence analysis of Epstein-Barr virus strain GD1 from a nasopharyngeal carcinoma patient. *J Virol* 79:15323–15330. <https://doi.org/10.1128/JVI.79.24.15323-15330.2005>.
 22. Liu P, Fang X, Feng Z, Guo YM, Peng RJ, Liu T, Huang Z, Feng Y, Sun X, Xiong Z, Guo X, Pang SS, Wang B, Lv X, Feng FT, Li DJ, Chen LZ, Feng QS, Huang WL, Zeng MS, Bei JX, Zhang Y, Zeng YX. 2011. Direct sequencing and characterization of a clinical isolate of Epstein-Barr virus from nasopharyngeal carcinoma tissue by using next-generation sequencing technology. *J Virol* 85:11291–11299. <https://doi.org/10.1128/JVI.00823-11>.
 23. Kwok H, Wu CW, Palser AL, Kellam P, Sham PC, Kwong DL, Chiang AK. 2014. Genomic diversity of Epstein-Barr virus genomes isolated from primary nasopharyngeal carcinoma biopsy samples. *J Virol* 88:10662–10672. <https://doi.org/10.1128/JVI.01665-14>.
 24. Kwok H, Tong AH, Lin CH, Lok S, Farrell PJ, Kwong DL, Chiang AK. 2012. Genomic sequencing and comparative analysis of Epstein-Barr virus genome isolated from primary nasopharyngeal carcinoma biopsy. *PLoS One* 7:e36939. <https://doi.org/10.1371/journal.pone.0036939>.
 25. Tso KK, Yip KY, Mak CK, Chung GT, Lee SD, Cheung ST, To KF, Lo KW. 2013. Complete genomic sequence of Epstein-Barr virus in nasopharyngeal carcinoma cell line C666-1. *Infect Agents Cancer* 8:29. <https://doi.org/10.1186/1750-9378-8-29>.
 26. Xiao K, Yu Z, Li X, Tang K, Tu C, Qi P, Liao Q, Chen P, Zeng Z, Li G, Xiong W. 2016. Genome-wide analysis of Epstein-Barr virus (EBV) integration and strain in C666-1 and Raji cells. *J Cancer* 7:214–224. <https://doi.org/10.7150/jca.13150>.
 27. Murata T, Kondo Y, Sugimoto A, Kawashima D, Saito S, Isomura H, Kanda T, Tsurumi T. 2012. Epigenetic histone modification of Epstein-Barr virus BZLF1 promoter during latency and reactivation in Raji cells. *J Virol* 86:4752–4761. <https://doi.org/10.1128/JVI.06768-11>.
 28. Lei H, Li T, Hung GC, Li B, Tsai S, Lo SC. 2013. Identification and characterization of EBV genomes in spontaneously immortalized human peripheral blood B lymphocytes by NGS technology. *BMC Genomics* 14:804. <https://doi.org/10.1186/1471-2164-14-804>.
 29. Liu Y, Yang W, Pan Y, Ji J, Lu Z, Ke Y. 2016. Genome-wide analysis of Epstein-Barr virus (EBV) isolated from EBV-associated gastric carcinoma (EBVaGC). *Oncotarget* 7:4903–4914. <https://doi.org/10.18632/oncotarget.6751>.
 30. Tarbouriech N, Buisson M, Geoui T, Daenke S, Cusack S, Burmeister WP. 2006. Structural genomics of the Epstein-Barr virus. *Acta Crystallogr D Biol Crystallogr* 62:1276–1285. <https://doi.org/10.1107/S0907444906030034>.
 31. Tang Y, He Y, Shi L, Yang L, Wang J, Lian Y, Fan C, Zhang P, Guo C, Zhang S, Gong Z, Li X, Xiong F, Li X, Li Y, Li G, Xiong W, Zeng Z. 2017. Coexpression of AFAP1-AS1 and PD-1 predicts poor prognosis in nasopharyngeal carcinoma. *Oncotarget* 8:39001–39011. <https://doi.org/10.18632/oncotarget.16545>.
 32. Wang J, Mei F, Gao X, Wang S. 2016. Identification of genes involved in Epstein-Barr virus-associated nasopharyngeal carcinoma. *Oncol Lett* 12:2375–2380.
 33. Ding L, Li L, Yang J, Zhou S, Li W, Tang M, Shi Y, Yi W, Cao Y. 2007. Latent membrane protein 1 encoded by Epstein-Barr virus induces telomerase activity via p16INK4A/Rb/E2F1 and JNK signaling pathways. *J Med Virol* 79:1153–1163. <https://doi.org/10.1002/jmv.20896>.
 34. Li L, Li Z, Zhou S, Xiao L, Guo L, Tao Y, Tang M, Shi Y, Li W, Yi W, Cao Y. 2007. Ubiquitination of MDM2 modulated by Epstein-Barr virus encoded latent membrane protein 1. *Virus Res* 130:275–280. <https://doi.org/10.1016/j.virusres.2007.05.013>.
 35. Zheng H, Li M, Ren W, Zeng L, Liu HD, Hu D, Deng X, Tang M, Shi Y, Gong J, Cao Y. 2007. Expression and secretion of immunoglobulin alpha heavy chain with diverse VDJ recombinations by human epithelial cancer cells. *Mol Immunol* 44:2221–2227. <https://doi.org/10.1016/j.molimm.2006.11.010>.
 36. Hu D, Zheng H, Liu H, Li M, Ren W, Liao W, Duan Z, Li L, Cao Y. 2008. Immunoglobulin expression and its biological significance in cancer cells. *Cell Mol Immunol* 5:319–324. <https://doi.org/10.1038/cmi.2008.39>.
 37. Luo C, Wang JJ, Li YH, Hu JY, Li GC. 2010. Immunogenicity and efficacy of a DNA vaccine encoding a human anti-idiotypic single chain antibody against nasopharyngeal carcinoma. *Vaccine* 28:2769–2774. <https://doi.org/10.1016/j.vaccine.2010.01.033>.
 38. Hu D, Duan Z, Li M, Jiang Y, Liu H, Zheng H, Li L, Bode AM, Dong Z, Cao Y. 2011. Heterogeneity of aberrant immunoglobulin expression in cancer cells. *Cell Mol Immunol* 8:479–485. <https://doi.org/10.1038/cmi.2011.25>.
 39. Yu Z, Lu J, Yu H, Yan Q, Zuo L, Li G. 2011. A precise excision of the complete Epstein-Barr virus genome in a plasmid based on a bacterial artificial chromosome. *J Virol Methods* 176:103–107. <https://doi.org/10.1016/j.jviromet.2011.06.015>.
 40. Li M, Zheng H, Duan Z, Liu H, Hu D, Bode A, Dong Z, Cao Y. 2012. Promotion of cell proliferation and inhibition of ADCC by cancerous immunoglobulin expressed in cancer cell lines. *Cell Mol Immunol* 9:54–61. <https://doi.org/10.1038/cmi.2011.40>.

41. Yu H, Lu J, Zuo L, Yan Q, Yu Z, Li X, Huang J, Zhao L, Tang H, Luo Z, Liao Q, Zeng Z, Zhang J, Li G. 2012. Epstein-Barr virus downregulates microRNA 203 through the oncoprotein latent membrane protein 1: a contribution to increased tumor incidence in epithelial cells. *J Virol* 86:3088–3099. <https://doi.org/10.1128/JVI.05901-11>.
42. Zheng Y, Zhang W, Ye Q, Zhou Y, Xiong W, He W, Deng M, Zhou M, Guo X, Chen P, Fan S, Liu X, Wang Z, Li X, Ma J, Li G. 2012. Inhibition of Epstein-Barr virus infection by lactoferrin. *J Innate Immunity* 4:387–398. <https://doi.org/10.1159/000336178>.
43. Duan Z, Zheng H, Xu S, Jiang Y, Liu H, Li M, Hu D, Li W, Bode AM, Dong Z, Cao Y. 2014. Activation of the Ig α 1 promoter by the transcription factor Ets-1 triggers Ig α 1-C α 1 germline transcription in epithelial cancer cells. *Cell Mol Immunol* 11:197–205. <https://doi.org/10.1038/cmi.2013.52>.
44. Duan Z, Zheng H, Liu H, Li M, Tang M, Weng X, Yi W, Bode AM, Cao Y. 2016. AID expression increased by TNF- α is associated with class switch recombination of Ig α gene in cancers. *Cell Mol Immunol* 13:484–491. <https://doi.org/10.1038/cmi.2015.26>.
45. Zeng Z, Fan S, Zhang X, Li S, Zhou M, Xiong W, Tan M, Zhang W, Li G. 2016. Epstein-Barr virus-encoded small RNA 1 (EBER-1) could predict good prognosis in nasopharyngeal carcinoma. *Clin Transl Oncol* 18: 206–211. <https://doi.org/10.1007/s12094-015-1354-3>.
46. Zeng ZY, Zhou YH, Zhang WL, Xiong W, Fan SQ, Li XL, Luo XM, Wu MH, Yang YX, Huang C, Cao L, Tang K, Qian J, Shen SR, Li GY. 2007. Gene expression profiling of nasopharyngeal carcinoma reveals the abnormally regulated Wnt signaling pathway. *Hum Pathol* 38:120–133. <https://doi.org/10.1016/j.humpath.2006.06.023>.
47. Zeng Z, Zhou Y, Xiong W, Luo X, Zhang W, Li X, Fan S, Cao L, Tang K, Wu M, Li G. 2007. Analysis of gene expression identifies candidate molecular markers in nasopharyngeal carcinoma using microdissection and cDNA microarray. *J Cancer Res Clin Oncol* 133:71–81. <https://doi.org/10.1007/s00432-006-0136-2>.
48. Yang Y, Zhou H, Yang Y, Li W, Zhou M, Zeng Z, Xiong W, Wu M, Huang H, Zhou Y, Peng C, Huang C, Li X, Li G. 2007. Lipopolysaccharide (LPS) regulates TLR4 signal transduction in nasopharynx epithelial cell line 5-8F via NF κ B and MAPKs signaling pathways. *Mol Immunol* 44: 984–992. <https://doi.org/10.1016/j.molimm.2006.03.013>.
49. Yang Y, Liao Q, Wei F, Li X, Zhang W, Fan S, Shi L, Li X, Gong Z, Ma J, Zhou M, Xiang J, Peng S, Xiang B, Deng H, Yang Y, Li Y, Xiong W, Zeng Z, Li G. 2013. LPLUNC1 inhibits nasopharyngeal carcinoma cell growth via down-regulation of the MAP kinase and cyclin D1/E2F pathways. *PLoS One* 8:e62869. <https://doi.org/10.1371/journal.pone.0062869>.
50. Zhang W, Huang C, Gong Z, Zhao Y, Tang K, Li X, Fan S, Shi L, Li X, Zhang P, Zhou Y, Huang D, Liang F, Zhang X, Wu M, Cao L, Wang J, Li Y, Xiong W, Zeng Z, Li G. 2013. Expression of LINC00312, a long intergenic noncoding RNA, is negatively correlated with tumor size but positively correlated with lymph node metastasis in nasopharyngeal carcinoma. *J Mol Histol* 44:545–554. <https://doi.org/10.1007/s10735-013-9503-x>.
51. Gong Z, Zhang S, Zeng Z, Wu H, Yang Q, Xiong F, Shi L, Yang J, Zhang W, Zhou Y, Zeng Y, Li X, Xiang B, Peng S, Zhou M, Li X, Tan M, Li Y, Xiong W, Li G. 2014. LOC401317, a p53-regulated long noncoding RNA, inhibits cell proliferation and induces apoptosis in the nasopharyngeal carcinoma cell line HNE2. *PLoS One* 9:e110674. <https://doi.org/10.1371/journal.pone.0110674>.
52. Liao Q, Zeng Z, Guo X, Li X, Wei F, Zhang W, Li X, Chen P, Liang F, Xiang B, Ma J, Wu M, Tang H, Deng M, Zeng X, Tang K, Xiong W, Li G. 2014. LPLUNC1 suppresses IL-6-induced nasopharyngeal carcinoma cell proliferation via inhibiting the Stat3 activation. *Oncogene* 33:2098–2109. <https://doi.org/10.1038/ncr.2013.161>.
53. Bo H, Gong Z, Zhang W, Li X, Zeng Y, Liao Q, Chen P, Shi L, Lian Y, Jing Y, Tang K, Li Z, Zhou Y, Zhou M, Xiang B, Li X, Yang J, Xiong W, Li G, Zeng Z. 2015. Upregulated long noncoding RNA AFAP1-AS1 expression is associated with progression and poor prognosis of nasopharyngeal carcinoma. *Oncotarget* 6:20404–20418. <https://doi.org/10.18632/oncotarget.4057>.
54. Zhang W, Fan S, Zou G, Shi L, Zeng Z, Ma J, Zhou Y, Li X, Zhang X, Li X, Tan M, Xiong W, Li G. 2015. Lactotransferrin could be a novel independent molecular prognosticator of nasopharyngeal carcinoma. *Tumour Biol* 36:675–683. <https://doi.org/10.1007/s13277-014-2650-1>.
55. Gong Z, Yang Q, Zeng Z, Zhang W, Li X, Zu X, Deng H, Chen P, Liao Q, Xiang B, Zhou M, Li X, Li Y, Xiong W, Li G. 2016. An integrative transcriptomic analysis reveals p53 regulated miRNA, mRNA, and lncRNA networks in nasopharyngeal carcinoma. *Tumour Biol* 37:3683–3695. <https://doi.org/10.1007/s13277-015-4156-x>.
56. Xu K, Xiong W, Zhou M, Wang H, Yang J, Li X, Chen P, Liao Q, Deng H, Li X, Li G, Zeng Z. 2016. Integrating ChIP-sequencing and digital gene expression profiling to identify BRD7 downstream genes and construct their regulating network. *Mol Cell Biochem* 411:57–71. <https://doi.org/10.1007/s11010-015-2568-y>.
57. Yang L, Tang Y, He Y, Wang Y, Lian Y, Xiong F, Shi L, Zhang S, Gong Z, Zhou Y, Liao Q, Zhou M, Li X, Xiong W, Li Y, Li G, Zeng Z, Guo C. 2017. High Expression of LINC01420 indicates an unfavorable prognosis and modulates cell migration and invasion in nasopharyngeal carcinoma. *J Cancer* 8:97–103. <https://doi.org/10.7150/jca.16819>.
58. Zhou Y, Liao Q, Li X, Wang H, Wei F, Chen J, Yang J, Zeng Z, Guo X, Chen P, Zhang W, Tang K, Li X, Xiong W, Li G. 2016. HYU1, regulated by LPLUNC1, is upregulated in nasopharyngeal carcinoma and associated with poor prognosis. *J Cancer* 7:367–376. <https://doi.org/10.7150/jca.13695>.
59. Tse KP, Su WH, Chang KP, Tsang NM, Yu CJ, Tang P, See LC, Hsueh C, Yang ML, Hao SP, Li HY, Wang MH, Liao LP, Chen LC, Lin SR, Jorgensen TJ, Chang YS, Shugart YY. 2009. Genome-wide association study reveals multiple nasopharyngeal carcinoma-associated loci within the HLA region at chromosome 6p21.3. *Am J Hum Genet* 85:194–203. <https://doi.org/10.1016/j.ajhg.2009.07.007>.
60. Feng BJ, Huang W, Shugart YY, Lee MK, Zhang F, Xia JC, Wang HY, Huang TB, Jian SW, Huang P, Feng QS, Huang LX, Yu XJ, Li D, Chen LZ, Jia WH, Fang Y, Huang HM, Zhu JL, Liu XM, Zhao Y, Liu WQ, Deng MQ, Hu WH, Wu SX, Mo HY, Hong MF, King MC, Chen Z, Zeng YX. 2002. Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. *Nat Genet* 31:395–399.
61. Hu LF, Qiu QH, Fu SM, Sun D, Magnusson K, He B, Lindblom A, Ernberg I. 2008. A genome-wide scan suggests a susceptibility locus on 5p13 for nasopharyngeal carcinoma. *Eur J Human Genet* 16:343–349. <https://doi.org/10.1038/sj.ejhg.5201951>.
62. Zeng Z, Huang H, Huang L, Sun M, Yan Q, Song Y, Wei F, Bo H, Gong Z, Zeng Y, Li Q, Zhang W, Li X, Xiang B, Li Y, Xiong W, Li G. 2014. Regulation network and expression profiles of Epstein-Barr virus-encoded microRNAs and their potential target host genes in nasopharyngeal carcinomas. *Sci China Life Sci* 57:315–326. <https://doi.org/10.1007/s11427-013-4577-y>.
63. Chen F, Liu J, Fan M, Wei XM, Xie YL, Wang LH, Yang H. 2014. Module function and two-way clustering analysis of Epstein-Barr virus-related nasopharyngeal cancer. *Genet Mol Res* 13:1823–1831. <https://doi.org/10.4238/2014.March.17.10>.
64. Edwards RH, Marquitz AR, Raab-Traub N. 2008. Epstein-Barr virus BART microRNAs are produced from a large intron prior to splicing. *J Virol* 82:9094–9106. <https://doi.org/10.1128/JVI.00785-08>.
65. Song Y, Li X, Zeng Z, Li Q, Gong Z, Liao Q, Chen P, Xiang B, Zhang W, Xiong F, Zhou Y, Zhou M, Ma J, Li Y, Chen X, Li G, Xiong W. 2016. Epstein-Barr virus encoded miR-BART11 promotes inflammation-induced carcinogenesis by targeting FOXPI. *Oncotarget* 7:36783–36799. <https://doi.org/10.18632/oncotarget.9170>.
66. Yan Q, Zeng Z, Gong Z, Zhang W, Li X, He B, Song Y, Li Q, Zeng Y, Liao Q, Chen P, Shi L, Fan S, Xiang B, Ma J, Zhou M, Li X, Yang J, Xiong W, Li G. 2015. EBV-miR-BART10-3p facilitates epithelial-mesenchymal transition and promotes metastasis of nasopharyngeal carcinoma by targeting BTRC. *Oncotarget* 6:41766–41782. <https://doi.org/10.18632/oncotarget.3258>.
67. He B, Li W, Wu Y, Wei F, Gong Z, Bo H, Wang Y, Li X, Xiang B, Guo C, Liao Q, Chen P, Zu X, Zhou M, Ma J, Li X, Li Y, Li G, Xiong W, Zeng Z. 2016. Epstein-Barr virus-encoded miR-BART6-3p inhibits cancer cell metastasis and invasion by targeting long noncoding RNA LOC553103. *Cell Death Dis* 7:e2353. <https://doi.org/10.1038/cddis.2016.253>.
68. Hua L, Xia H, Zhou P, Li D, Li L. 2014. Combination of microRNA expression profiling with genome-wide SNP genotyping to construct a coronary artery disease-related miRNA-miRNA synergistic network. *Bio-sci Trends* 8:297–307. <https://doi.org/10.5582/bst.2014.01031>.
69. Yu J, Liu Y, Guo C, Zhang S, Gong Z, Tang Y, Yang L, He Y, Lian Y, Li X, Deng H, Liao Q, Li X, Li Y, Li G, Zeng Z, Xiong W, Yang X. 2017. Upregulated long noncoding RNA LINC00152 expression is associated with progression and poor prognosis of tongue squamous cell carcinoma. *J Cancer* 8:523–530. <https://doi.org/10.7150/jca.17510>.
70. Yu J, Liu Y, Gong Z, Zhang S, Guo C, Li X, Tang Y, Yang L, He Y, Wei F, Wang Y, Liao Q, Zhang W, Li X, Li Y, Li G, Xiong W, Zeng Z. 2016. Overexpression long noncoding RNA LINC00673 is associated with poor prognosis and promotes invasion and metastasis in tongue squamous

- cell carcinoma. *Oncotarget* 8:16621–16632. <https://doi.org/10.18632/oncotarget.14200>.
71. Zhao R, Liu Y, Wang H, Yang J, Niu W, Fan S, Xiong W, Ma J, Li X, Phillips JB, Tan M, Qiu Y, Li G, Zhou M. 2016. BRD7 plays an anti-inflammatory role during early acute inflammation by inhibiting activation of the NF- κ B signaling pathway. *Cell Mol Immunol* <https://doi.org/10.1038/cmi.2016.31>.
 72. Zeng Z, Bo H, Gong Z, Lian Y, Li X, Li X, Zhang W, Deng H, Zhou M, Peng S, Li G, Xiong W. 2016. AFAP1-AS1, a long noncoding RNA upregulated in lung cancer and promotes invasion and metastasis. *Tumour Biol* 37:729–737. <https://doi.org/10.1007/s13277-015-3860-x>.
 73. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18. <https://doi.org/10.1186/2047-217X-1-18>.
 74. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
 75. Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286–298. <https://doi.org/10.1093/bib/bbn013>.
 76. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739. <https://doi.org/10.1093/molbev/msr121>.